First keep a copy for all the old data and we will make modifications based on it

1. Eigenvector

* Per-Chromosome Median Calculation (Per Sample)

Get the median value of both (1) positive eigen vector and (2) negative eigen vector from each chromosome, in total 23 chromosomes

For compartment A,

chr1 median value of positive eigen vector is A

chr2 median value of positive eigen vector is A’

For each chromosome (chr1 to chr22, chrX): Compute the median of positive eigenvector values (A, A', A'', ...)

For compartment B,

chr1 median value of negative eigen vector is B

chr2 median value of negative eigen vector is B’

For each chromosome (chr1 to chr22, chrX): Compute the median of negative eigenvector values (B, B', B'', ...)

* Normalization **to make sure data across chromosome and across samples are all comparable:**

A math equations on a white background

AI-generated content may be incorrect.

Scale down each sample to make sure data across chromosome and across samples are all comparable: X(A) = mean of the median values = ( A + A’ + A’’ + A’’’ + …)/n

Eg. scale down chr1 in sample 1 by A/X(A)

a1 ÷ A/X(A)

a2 ÷ A/X(A)

a3 ÷ A/X(A)

X(B) = mean of the median value = ( B + B’ + B’’ + B’’’ + …)/n

Eg. scare down chr1 in sample 1 by B/X(B)

b1 ÷ B/X(B)

b2 ÷ B/X(B)

**After adjust EV values we then process to MA plot**

1. MA plot

* We will focus on non-symmetric data points and this might be caused by EBV infection
* Histogram comes from MA plot, for histogram we need to make sure the axis are of same length
* From MA plot we will know the noise range, and we will need to make a cut off that screen eg.2k genes, and then short it down to several key genes
* **For this MA plot, we will plot EV from all chromosomes of 6 samples together in one plot (I did not found such MA plot in Xiang’s NC paper)**

Steps to Generate MA Plot

1. Prepare normalized eigenvector data

2. MA Plot – Each Sample vs Mean of All Samples

M = ?

A = average of the EV values across 6 samples

3. Compute M and A values

###### Xiang’s script for MA plot ############

# MA plot

pdf('ev.ma.pdf',9,69)

par(mfrow=c(23,3),font.lab=2,cex.lab=1.2)

for(i in seq\_along(EV.rbl)){

EV\_diff <- EV.lcl[[i]]-EV.rbl[[i]]

EV\_mean <- (EV.rbl[[i]]+EV.lcl[[i]])/2

plot(EV\_mean,EV\_diff,pch=16,main=paste0(names(EV.rbl)[i]),cex=0.25)

abline(h=0,lty="dashed")

}

dev.off()

* **Histogram we need to refer to Xiang, where he draw the histogram of whole genome**

Reference

A comparison of a normal distribution graph

AI-generated content may be incorrect.